Examination

Linköping University, Department of Computer and Information Science, Statistics

Course code and name	732A38 Computational Statistics
Date and time	2016-03-23, 8.00-13.00
Assisting teacher	Oleg Sysoev
Allowed aids	"Computational statistics" by Gentle or "Computational Statistics" by Givens &Hoeting
Grades:	A=19-20 points B=16-18 points C=11-15 points D=9-10 points E=7-8 points F=0-6 points

Provide a detailed report that includes plots, conclusions and interpretations. Give motivated answers to the questions. If an answer is not motivated, the points are reduced. Provide all necessary codes in the appendix.

Assignment 1 (10p)

1. Implement a function that generates a random number from the probability distribution given by $f(x) = 1.5\sqrt{x}, 0 < x < 1$. Generate a sample containing 1000 random numbers from this distribution, plot the histogram of this sample and comment whether the sample looks like it comes from $f(x)$ **(3p)**
2. Implement a Metropolis-Hastings algorithm that simulates from $g(x) \propto \frac{\sqrt{x}}{x+0.1}, 0 < x < 1$ by using beta distribution with $shape2 = 0.5$ as proposal distribution. Generate 1000 random numbers from $g(x)$ by using starting point $x_0 = 0.1$, make a trace plot and comment on the burn-in period, mixing and convergence. **(4p)**
3. Estimate $\int_0^1 \frac{x\sqrt{x}}{x+0.1}$ by using a) samples obtained from step 1; b) samples obtained from step 2. Compute the true value of the integral by using integrate() and comment which estimator

appeared to be the best. Recall the properties of the importance sampling and MCMC sampling and comment why one estimation method appeared to be better than another one. **(3p)**

Assignment 2 (10p)

Data file "**process.csv**" describes the outcome Y of a detector based on its power measurement X

1. Fit a LOESS smoother to the data, present Y versus X and also the fitted values versus X in one plot. Does it look like a quadratic polynomial can describe the probability of the outcome well? **(1p)**
2. Assume that the data comes from the model
$$Y_i \sim Bernoulli(p_i), p_i = \max\left(0.1, min\left(0.9, w_1 X_i + w_2 X_i^2\right)\right)$$
where the probability mass function of the Bernoulli distribution is
$$p(Y_i) = p_i^{Y_i}(1-p_i)^{1-Y_i}$$
and w_1 and w_2 are unknown parameters.
Augment your data with a new variable $Z = X^2$ and derive a minus log-likelihood expression for the model described above (assume Y_i being independent). Afterwards, obtain optimal parameters for w_1 and w_2 by using BFGS method with starting points
    a) $w_1 = 0.1$ and $w_2 = 0.1$
    b) $w_1 = 0.3$ and $w_2 = 0.3$
    c) $w_2 = 0.9$ and $w_2 = 0.1$

    Comment on the optimal parameter values and convergence in each case. **(4p)**
3. Present the data and the fitted function $\hat{Y} = \max(0.01, min(0.99, w_1 X + w_2 X^2))$ in the same plot, where $\hat{w}_1$ and $\hat{w}_2$ are the optimal estimates you found in step 2a. Compare this result to the LOESS fit. **(1p)**
4. Implement a permutation test with $B = 200$ iterations that uses test statistics $T = \hat{w}_2$ (T is obtained by applying our optimizer with starting point $w_1 = 0.1$ and $w_2 = 0.1$ to the current data and then the estimated w_2 is returned as T) to test $H_0: w_2 = 0$ vs $H_a: w_2 \neq 0$. Run this test, plot the histogram of the test statistics and present your conclusions. What is the p-value of this test? **(4p)**
    - In case you didn't succeed with steps 1-3, you may use a linear regression model to get the estimates of $\hat{w}_1$ and $\hat{w}_2$ in the permutation test, but the points will be reduced